Chapter 5

Cloning of N terminal domain of *Anabaena* sp. strain PCC 7120 XisA protein and investigation of recombinase activity

5.1 Introduction

Site-specific recombination process is responsible for integration and excision of bacteriophage genomes into and out of their bacterial host chromosome. Integrases are a family of proteins that recombine DNA duplexes by executing two consecutive strand breakage and rejoining steps and a topoisomerisation of the reactants (Robert *et al.*,2002). The first member of this family, the well-studied lambda Integrase protein, promotes integration and excision of the λ phage genome from that of the host (Landy, 1989); other family members function in the maintenance of plasmid copy number (Abremski and Hosess, 1984; Sadowski *et al.*,1987), the elimination of dimers from replicated chromosomes (Hayes and Sherratt, 1997), alteration of cell-surface components (McClain *et al.*,1991) and in the life cycle of temperate phages (Waldman *et al.*,1986; Yu *et al.*,1989; Lee and Hatfull, 1993). All these processes involve the conservative site-specific recombination of two DNA target sites.

Bacteriophage lambda Integrase (Int) is a recombinase which inserts and excises the phage genome into and out of the *Escherichia coli* chromosome. It belongs to a large family of tyrosine recombinases (more than 300) whose members carry out site-specific recombination of phage, plasmid, and chromosomal sequences (Esposito and Scocca, 1997; Nunes-Duby *et al.*,1998; Groth and Calos, 2004). Int mediates recombination via four distinct pathways (summarized in Table 1), in which different protein-DNA complexes are assembled by Int and its accessory factors on four types of recombination target sites, known as attachment or *att* sites. In addition to the catalytic domain, Int has two DNA binding domains: a relatively high-affinity binding domain contacts the "arm" binding sites, which

are distal to the loci of strand exchange, while a low-affinity binding domain contacts the "core" binding sites, which directly flank the loci of strand exchange (Ross and Landy, 1982; Tirumalai *et al.*,1996). Int cleaves and rejoins four phosphodiester bonds in the host and phage DNAs in two stages involving cleavage and ligation occurring at one strand at each target site. It involves transient covalent intermediates of tyrosine phosphodiester bond and a free 5' hydroxyl group which gets ligated with other target site intermediate. Similar events occurring at the second strand separates four stranded intermediate.

Presumably because abortive reactions would be detrimental both to the phage and to the host, the recombinases have evolved to be very efficient, and very few abortive products are seen (Pargellis *et al.*, 1988). At worst, aborted recombination events would generate breaks in both the bacterial and phage genomes; at best, integration or excision would be less efficient due to unproductive rounds of strand cleavage and joining and the lysogeny versus lysis decision of the phage would be impeded. Abortive recombination may be minimized by carrying out strand exchange in the context of specific synaptic complexes containing both of the DNA substrates and all of the protein subunits necessary for the reaction. This indeed appears to be the case (Richet *et al.*,1988; Segall and Nash, 1993; Burgin and Nash, 1995; Segall, 1998). Moreover, the DNA strand cleavage and ligation events in integrative and excisive recombination are highly concerted, for which interactions between Int monomers are presumed to be important. Protein-protein interactions appear to be a central feature of site-specific recombination reactions, including phase variation of flagellar antigens in *Salmonella enterica* (Lim, 1994; Haykinson *et al.*,1996), resolution of γδ transposon generated cointegrate structures (Hughes *et al.*,1993; Murley and Grindley, 1998), and 2μ plasmid inversion mediated by Flp (Lee *et al.*,1996 and references therein).

The DNA binding and catalytic properties of Int, as well as those of many protein-DNA intermediates assembled by Int in conjunction with its accessory factors, have been studied extensively (reviewed in Landy, 1989; Nash, 1996). However, no direct information is available on the interactions between Int monomers occurring during recombination. Two major observations suggest that protein-protein contacts between Int monomers are important. First, binding of Int to the *att* sites is highly cooperative (Kim *et al.*,1990; Segall

and Nash, 1993). Second, different catalytic mutants of the Int protein complement each other for strand cleavage activity (Han *et al.*,1994), suggesting intimate interactions between at least two Int monomers. More recently, further information has become available from several crystal structures of Int family proteins. Four structures have been solved, namely, those of the catalytic domain of Int (Kwon *et al.*,1997), the catalytic domain of the *Haemophilus influenzae* phage HP1 Int (Hickman, 1997), the *E. coli* XerD protein (Subramanya *et al.*,1997), and the cocrystal of the bacteriophage P1 Cre protein bound to a Holliday junction (Gopaul *et al.*,1998; Guo *et al.*,1997). The Int structure and the XerD structure were solved as monomers and thus provide information on protein-protein contacts only by homology alignments with the catalytic domains of HP1 Int, solved as a dimer, and with the Cre structure, which was solved as a pseudosymmetric tetramer (Guo *et al.*,1997). However, Int and the other bacteriophage integrases share little or no recognizable homology in the noncatalytic domains of their relatives (Esposito and Scocca, 1997).

The initial definition of the family was based on comparisons of seven sequences, and three invariant residues were identified: a His–X–X–Arg cluster and a Tyr residue (Argos *et al.*, 1986). Alignment of 28 sequences identified a fourth invariant position, occupied by an Arg residue (Abrenski and Hoess, 1992). These four conserved residues are located in the C-terminal half of the protein sequences, and occur in the order Arg, His–X–X–Arg and Tyr, with Tyr closest to the C-terminus. Mutations introduced at each of these conserved or invariant positions in several different systems produced proteins inactive in recombination, as would be expected if these positions corresponded to active site residues (Parson *et al.*, 1988; Friesen and Sadowski, 1992; Abrenski and Hoess, 1992; Han *et al.*, 1994; Groth and Calos, 2004).

Site specific DNA recombination reactions are catalyzed by polynucleotidyl transferases that are specific to each reaction system (Mizuuchi, 1997). Two major families of site-specific recombinases (tyrosine and serine recombinases) use different mechanisms for cutting and rejoining the DNA strands at the recombination crossover sites. The tyrosine recombinases complete the cleavage, exchange and rejoining of one pair of DNA strands with the generation of a Holliday junction as a recombination intermediate, before initiation

the same set of reactions on the other pair of DNA strands. If the recombination sites are on different DNA molecules, the net result of the process is the integration of the two molecules into a single one. However, if the recombination sites are on the same DNA molecule, the recombination reaction can lead to either the deletion or the inversion of the DNA sequences located between the crossover sites, depending on their relative orientation and number of inter domanial super coils trapped on syanapsis (Nash, 1996; Hallet and Sherratt, 1997; Grindley, 2002). Two subfamilies of serine recombinases are known. One includes small recombinases (<250 amino acids long) that primarily catalyze intramolecular recombination reactions (Grindley, 2002). Second subfamily comprises large recombinases (>400 amino acids long) that catalyze both inter- and intramolecular recombination events (Smith and Thorpe, 2002).

XisA protein has similarity with DNA breaking-joining enzyme super family which includes type IB topoisomerases and tyrosine recombinases that share the same fold in their catalytic domain containing six active site residues. Endonuclease activity is only possible when enzyme cuts at both the strand of DNA which is characteristic of the serine class of resolvases rather than sequential single stranded break followed by ligation that is characteristics of tyrosine class of integrases. Endonuclease type II e and type II f cleaves DNA at specific sites, which have dyad symmetry but require two such sites which can be effective even in *trans* (Jo and Topal, 1995). Since C terminal of XisA has similarity with int family enzymes it is suspected that may be N terminal of the XisA may have endonuclease activity. In order to determine the functional distribution of the activity of XisA, the protein was divided in to N terminal and C terminal domain. Both the domains were monitored for rearrangement.

Summary of four pathways of bacteriophage λ site specific recombination (Jessop *et al.*, 2000).

Component(s)	**Pathway**			
	Integration	**Excision**	**Bent-L**	**Straight-L**
att substrates	*attP, attB*	*attL, attR*	*attL(tenP'I)*	*attL*
int requirement	Y	Y	Y	Y
Bending protein requirement	IHF	IHF or HU	IHF	Inhibitory
Xis requirement	Inhibitory	Y	N	N
Supercoiling requirement	Y	N	N	N
Efficiency	High	High	High	Low

5.2 Materials and Methods

Please refer chapter titled Material and Methods

5.3 Results and Discussion

N terminal domain of *xisA* gene along with its promoter was amplified using specific primers (**Fig1**). N terminal domain was initially cloned in pTZ57R by TA cloning method and cloned in pNTMCGm plasmid (**Fig 2**) and the presence of the insert was confirmed by restriction endonuclease digestion (**Fig. 3**). *E. coli* strain DH5α harboring plasmid pMX32 and pNTMCGm was confirmed by the restriction digestion pattern of plasmids from *E. coli* transformants (**Fig 4**). It encodes for 287 amino acids residues of the N terminal domain ending with the arginine of RYRY tetrad (**Fig. 5**).

No rearrangement has been found with the N terminal domain in LB and M9 media while in case of positive control of substrate plasmid showed rearrangement of 0.58% in LB and 0.80% in M9 minimal media, respectively. To monitor the expression of the N terminal of *xisA* protein profiling was done using 15% SDS-PAGE. A protein band at about 33 kDa corresponding to the estimated size of N terminal domain was detected (**Fig. 6**).

Further purification is required for confirming the protein. These results demonstrate that N terminal domain is not catalytically active and supports the catalytic domain is C terminal domain. This is in agreement with the evidence demonstrating the functionality of C terminal integrase domain of XisC (Carrasco *et al.*, 2005) another homologous recombinase of *Anabaena* PCC 7120. Site-directed mutagenesis of tyrosine corresponding to RKHRY(W) catalytic site led to the loss of site specific recombinase activity (Groth and Calos, 2004). The 356 amino acid λ Integrase is divided into two domains by limited proteolysis. The N terminal domain includes residues 1-64 and is responsible for the binding of the arm types of attP **Fig. A** (Groth and Calos, 2004).

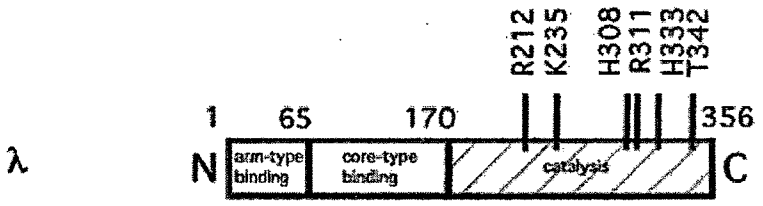

Fig. A Protein domain structure of λ integrase.

The C-terminal domain binds the lower affinity core-type sites and contains the catalytic site (Tirumalai *et al,* 1997). The C terminal activities can be further separated. Amino acid residues 65-169 are responsible for the core-type binding, while amino acid residues 170-356 constitue the minimal catalytic domain. (Tirumalai *et al.,* 1998). The catalytic tyrosine and the RKHRH pentad are contained in this region. Both the N-terminal and C-terminal domains contribute to protein-protein interactions between Integrase monomers. In the serine recombinase family, the smaller, better-studied resolvase and invertase members, having, 200 amino acid residues, have two distinct domains. An, 140 amino acid N-terminal domain is responsible for protein-protein interactions and catalysis, while a smaller C-terminal domain contains a helix-turn-helix motif and the DNA-binding function of the proteins (Abdel-Meguid *et al.,*1984; Newman and Grindley, 1984). The serine integrases, however, belong to a subfamily of much larger proteins (441-772 amino acid residues) and are similar to the resolvases/invertases predominantly in the catalytic N-terminal region and a stretch of, 220 amino acid residues immediately following it (Dyson and Farrell, 1985; Christiansen *et al.,* 1994; Thorpe and Smith, 2002). The additional amino acid residues found in the C-terminal domains contain little similarity to their resolvase counterparts. On the basis of sequence similarity and a SpoIVCA truncation mutant, it is believed that the DNA-binding function of ΦC31integrase resides in the N-terminal half of the protein (Popham and Stragier, 1992).

N terminal of *xisA* was amplified as shown in **Fig 1.**

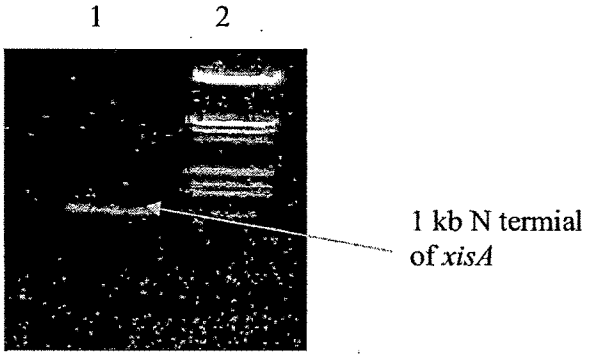

Fig. 1 Amplification of the N terminal domain of the *xisA* gene. Lane 1 Amplified N terminal domain of *xisA*, Lane 2. Lambda DNA cut with EcoR1/HindIII.

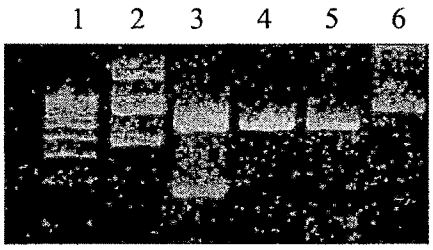

Fig. 2 Restriction digestion of T vector containing N terminal of *xisA* gene. Lane 1 500bp ladder, Lane2 Control undigested plasmid, Lane 3 pTZ57RNT digested with BclI, Lane 4 pTZ57R digested with HindIII, Lane 5 pTZ57RNT digested with HindIII and Lane 6 pTZ57RNT digested with XbaI.

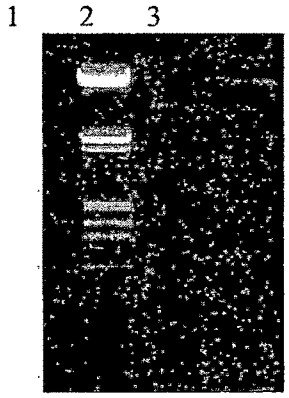

Fig. 3 Restriction digestion of the pNTMCGm plasmid. Lane 1 Lambda DNA cut with EcoR1/HindIII, Lane 2. pNTMCGm digested with EcoR1, Lane 3. pNTMCGm control plasmid.

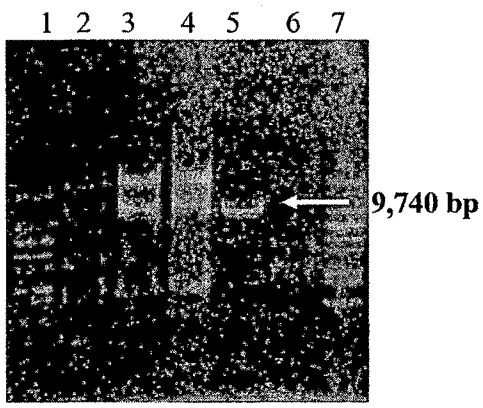

Fig. 4 Restriction digestion of *E. coli* DH5α transformants. Lane 1. λ DNA digested with BstEII, Lane 2. pMX32 digested with EcoR1, Lane 3. EcoRI digest of the plasmids of Colony 1 containing pMX32 and pNTMCGm plasmids, Lane 4. EcoRI digest of the plasmids of Colony 2 containing pMX32 and pNTMCGm plasmids, Lane 5. pNTMCGm digested with EcoR1, Lane 6. pNTMCGm undigested, Lane 7. Lambda DNA digested with BstEII.

1
MQNQGQDKYQ QAFADLEPLS STDGSFLGSS LQAQQQREHM RTKVLQDLDK
VNLRSKSAKT KVSVRESNGS LQLRATLPIK PRDKDTNGTG RKQYNLSLNI
PANLDGLKTA EEEAYELGKL IARKTFEWND KYLGKEATKK DSQTIGDLLE
KFAEEYFKTH KRTTKNKHTF FYYFSRTQRY TNSKDLATAE NLINSIEQID
KE'**WARYN**AAR AISAFCITFN IEIDLSQYSK MPDRNSRNIP TDAEILSGIT
KFEDYLVTRG NQVNEDVKDS WQLWRWTYGM LAVFGL**R**
287

Fig. 5 Amino acid sequence encoded by N terminal of *xisA* gene in pTZ57R and pNTMCGm plasmid.

Table 1 Endonuclease Assay for N-terminus

Conditions/Time **24 h**	**No. of colonies of strains DH5α grown in LB** **Mean CFU/ml**	**No. of colonies of strains DH5α grown in M9** **Mean CFU/ml**
	pMX32, pNTMC71AGm	pMX32, pNTMC71AGm
LA + X- gal	$2x10^{09}$±8.78 $x10^{08}$	$1.34x10^{09}$±6.35 $x10^{08}$
LA + G + X- gal	$1.7\ x10^{09}$±6.82 $x10^{08}$	$1.23\ x10^{09}$±1.01 $x10^{09}$
LA + A + X- gal	$1.5\ x10^{09}$±6.49 $x10^{08}$	$8.1\ x10^{08}$±7.58 $x10^{08}$
LA + A + G + X- gal	$1.3\ x10^{09}$±5.84 $x10^{08}$	$8.82\ x10^{08}$±8.63 $x10^{08}$

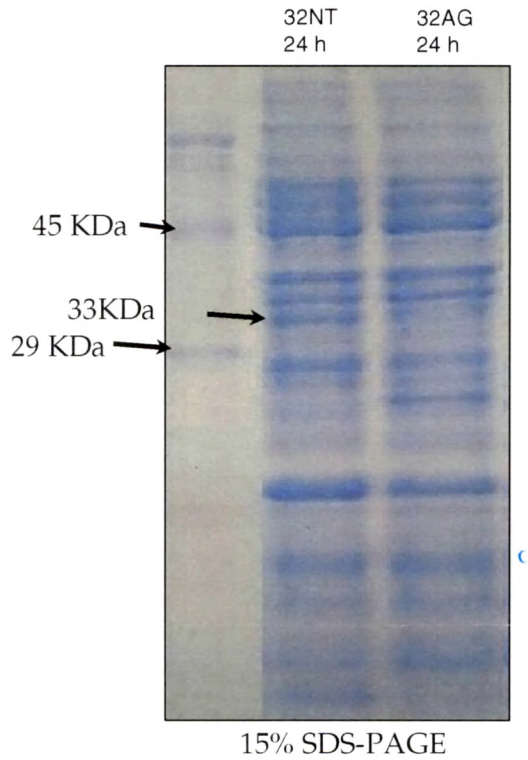

Fig. 6 Expression of N-terminus Protein in *E. coli* DH5α in LB.